INFORMATION AND CONTROL 26, 280–285 (1974)

# Parallel Context-Free Languages

SVEN SKYUM

*Department of Computer Science, University of Aarhus, Denmark*

*Communicated by Arto Salomaa*

The relation between the family of context-free languages and the family of parallel context-free languages is examined in this paper.

It is proved that the families are incomparable.

Finally we prove that the family of languages of finite index is contained in the family of parallel context-free languages.

## INTRODUCTION

It has been an open problem whether or not the family of context-free languages is contained in the family of parallel context-free languages (e.g., cf. Rosenfeld, 1971). By "parallel" we mean that whenever you use a production $A \rightarrow \alpha$ in a derivation you have to use this specific production simultaneously for all occurrences of $A$ in the sentential form you are dealing with.

In Siromoney and Krithivasan (1974) it is claimed that the intersection of the two families consists of languages of finite index.

The reason for writing this paper is that the proofs in Siromoney and Krithivasan (1974) contain so many serious gaps that it seems appropriate to find alternative proofs.

In this paper it is proved that the family of parallel context-free languages contains the family of languages of finite index but not the family of context-free languages.

It is assumed that the reader is familiar with the basic notions concerning formal language theory. For unexplained notions we refer to Salomaa (1973).

## The Relation Between Parallel Context-Free Languages and Context-Free Languages

Definition. Let $G = (V, T, P, S)$ be an ordinary context-free grammar. The parallel direct yield relation $\Rightarrow_G^{(P)}$ on the set $(V \cup T)^*$ is defined as follows:

$$\alpha \overset{(p)}{\underset{G}{\Rightarrow}} \beta \quad \text{iff} \quad \alpha = \alpha_1 A \alpha_2 A \cdots A \alpha_k \quad \text{and} \quad \beta = \alpha_1 w \alpha_2 w \cdots w \alpha_k ,$$

where $\alpha_i \in ((V \cup T)\backslash\{A\})^*$ for $1 \leqslant i \leqslant k$ and $A \to w$ is a production in $G$. The relation $\Rightarrow_G^{(P)*}$ is the reflexive transitive closure of $\Rightarrow_G^{(P)}$. $\alpha \Rightarrow_G^{(P)t} \beta$ iff there exists words $\alpha_i \in (V \cup T)^*$ for $0 \leqslant i \leqslant t$ such that

$$\alpha = \alpha_0 \overset{(p)}{\underset{G}{\Rightarrow}} \alpha_1 \overset{(p)}{\underset{G}{\Rightarrow}} \alpha_2 \overset{(p)}{\underset{G}{\Rightarrow}} \cdots \overset{(p)}{\underset{G}{\Rightarrow}} \alpha_t = \beta.$$

The language generated in parallel by $G$ is defined by:

$$L_p(G) = \{w \in T^* \mid S \overset{(p)*}{\underset{G}{\Rightarrow}} w\}$$

The family of parallel context-free languages is the family of languages generated in parallel by ordinary context-free grammars. The following example shows that the family of parallel context-free languages contains languages which are non-context-free.

Example. The language $\{a^{2^n} \mid n \geqslant 0\}$ is generated in parallel by the following grammar. $(\{S\}, \{a\}, \{S \to SS, S \to a\}, S)$.

Theorem 1. *The family of context-free languages is not contained in the family of parallel context-free languages.*

*Proof.* Let $L$ be the context-free language generated by the following grammar $G$.
$G = (\{S, A\}, \{0, 1, (,), [,]\}, P, S)$ where $P$ consists of the following productions.

$$S \to SS,\ S \to (0A1],\ S \to \lambda, \qquad A \to 0A1,\ A \to )S[.$$

$L$ is very similar to the Dyck-set (over 0 and 1), but in $L$ you have that corresponding groups of 0's and 1's are surrounded by parentheses and brackets, respectively.

Assume that there exists a grammar

$$H = (V, T, P, S) \text{ such that } H \text{ generates } L \text{ in parallel.}$$

Let $t$ be the number of nonterminals in $H$ and $m$ the length of the longest right side of a production in $H$. Let $\bar{m} = \lceil \log_2 m \rceil + 1$ (the least integer which is greater than or equal to $\log_2 m + 1$). We assume that no nonterminal is useless in $H$ and whenever nothing else is stated $\Rightarrow$ and $\Rightarrow^*$ are used instead of $\Rightarrow_H^{(P)}$ and $\Rightarrow_H^{(P)*}$.

LEMMA 1. *If $A \in V$ and $A \Rightarrow^* w_1 w_2$ then*

$$\#_1(w_1) \leqslant \#_0(w_1) + m^{2t} \quad \textit{and} \quad \#_0(w_2) \leqslant \#_1(w_2) + m^{2t}.$$

*Proof.* Since $A$ is not useless there exists a derivation of the form

$$S \stackrel{t'}{\Rightarrow} \alpha_1 A \alpha_2 \stackrel{t''}{\Rightarrow} v_1 A v_2 A \cdots A v_k \stackrel{*}{\Rightarrow} v_1 w_1 w_2 v_2 \cdots w_1 w_2 v_k \in L,$$

where $t', t'' \leqslant t$, $\alpha_i \in (V \cup T)^*$ for $1 \leqslant i \leqslant 2$, $v_i \in T^*$ for $1 \leqslant i \leqslant k$, $w_1 w_2 \in T^*$, and $A \Rightarrow^* w_1 w_2$. Therefore $|v_i| \leqslant m^{2t}$ for $1 \leqslant i \leqslant k$. As for the Dyck set (over 0 and 1) we have that in any prefix of a word in the language the number of 1's in the prefix is at most the number of 0's.

We have $\#_1(v_1 w_1) \leqslant \#_0(v_1 w_1)$ which implies that

$$\#_1(w_1) \leqslant \#_0(w_1) + |v_1| \leqslant \#_0 + m^{2t}.$$

The same kind of argument used on the suffix $w_2 v_k$ gives that

$$\#_0(w_2) \leqslant \#_1(w_2) + m^{2t}.$$

Define inductively $Q_i$ to be the languages.

$$Q_1 = \{(0^n)[1^n] \mid n \geqslant 1\}, \quad \text{and for} \quad i \geqslant 1,$$
$$Q_{i+1} = \{(0^n)\, w_1 w_2 [1^n] \mid n \geqslant 1;\, w_1, w_2 \in Q_i\}.$$

We have $Q_i \subseteq L$ for all $i \geqslant 1$.

Let $w \in Q_q$ be the word

$$(0^{n_1})(0^{n_2}) \cdots (0^{n_q})[1^{n_q}](0^{n_{q+1}}) \cdots [1^{n_2}](0^{n_{2^{q-1}+1}})(0^{n_{2^{q-1}+2}}) \cdots [1^{n_{2^{q-1}+1}}][1^{n_1}],$$

where $q = (t+2)\bar{m}$, $n_1 = m^{2t} + 1$, and $n_i = n_{i-1} + 1$, $2 \leqslant i \leqslant 2^q - 1$. $w$ will be fixed in the following and let

$$(1)\; S = w_0 \Rightarrow w_1 \Rightarrow w_2 \Rightarrow \cdots \Rightarrow w_n = w$$

be a fixed derivation of $w$ in $H$.

Lemma 2. *If $A$ occurs in $w_i$ for some $0 \leqslant i \leqslant n$ and this occurrence of $A$ derives the subword $w_A$ of $w$ in the derivation* (1), *then there exists integers $-2 \leqslant j \leqslant 0$ and $0 \leqslant k$ such that $\#,\ (w_A) = 2^k + j$. We will say that $A$ in that case is of degree $k$, if $k \geqslant 2$.*

*Proof.* Let $w_A$ be as above and let $h : T^* \to \{a, b\}^*$ be the homomorphism: $h()) = a$, $h([) = b$, and $h(0) = h(1) = h(() = h(]) = \lambda$. Let $v_i \in \{a, b\}^*$ be defined inductively by:

$$v_0 = \lambda$$
$$v_{i+1} = av_iv_ib, \qquad \text{for} \quad i \geqslant 0.$$

We then have that $h(w) = v_q$, and $\#_a(v_k) = 2^k - 1$. If $h(w) = x\underline{a}v_iv_i\underline{b}y$ for some $i$ and $x, y \in \{a, b\}^*$, then we say that the underlined occurrences of $a$ and $b$ correspond to each other, and we will use expressions as "the occurrence of $b$ corresponding to the occurrence of $a$" in this case.
Now if $|\, h(w_A)| \leqslant 1$ then $0 \leqslant \#_a(h(w_A)) = \#_)(w_A) \leqslant 1$ which is of the form $2^k + j$ for some $k \geqslant 0$ and $-2 \leqslant j \leqslant 0$.
Assume then that $|\, h(w_A)| > 1$.
There are then 4 cases to consider, namely:

(1) $h(w_A) = ava$ for some $v \in \{a, b\}^*$,

(2) $h(w_A) = avb$ for some $v \in \{a, b\}^*$,

(3) $h(w_A) = bva$ for some $v \in \{a, b\}^*$,

(4) $h(w_A) = bvb$ for some $v \in \{a, b\}^*$.

Cases (1), (3), and (4) are all treated in the same way. Assume, e.g., that $h(w_A) = ava$ for some $v \in \{a, b\}^*$, then $w_A = x(0^p)z$ where $\#_1(z) = 0$ and $p > m^{2t}$ which is contradictory to Lemma 1.

Assume then that $h(w_A) = avb$ for some $v \in \{a, b\}^*$. There are again more possibilities:

(1) The occurrence of $b$ in $h(w)$ which corresponds to the first occurrence of $a$ in the subword $h(w_A)$ is outside $h(w_A)$. The situation is then that the last occurrence of $b$ in $h(w_A)$ must correspond to an occurrence of a inside $h(w_A)$. We then have that $h(w_A) = av'av_iv_ib$ for some $0 \leqslant i \leqslant q$ and $v' \in \{a, b\}^*$.

(1a) $v' = v''a$.

Then $w_A = x(0^{p_1})(0^{p_2})\, y_1\, y_2 z$ for some $x, z \in T^*$ and $y_1, y_2 \in Q_i$.

$$\#_0((0^{p_1})(0^{p_2})\, y_1 y_2 z) - \#_1((0^{p_1})(0^{p_2})\, y_1 y_2 z)$$
$$= p_1 + p_2 + \#_0(z) - \#_1(z) \geqslant p_1 + \#_0(z) > m^{2t}$$

which is contradictory to Lemma 1.

(1b) $v' = v''b$. Then $h(w_A) = av_iv_ibav_iv_ib$ and $\#_)(w_A) = 2^{i+2} - 2$ or $h(w_A) = a\bar{v}av_iv_ibav_iv_ib$ for some $0 \leqslant i \leqslant q$ and $\bar{v} \in \{a, b\}^*$

(1ba) $\bar{v} = \bar{\bar{v}}a$. The situation is then as in (1a).

(1bb) $\bar{v} = \bar{\bar{v}}b$. This can not be the case because $bav_iv_ibav_iv_ib$ can not be a subword of $h(w)$.

(1bc) $\bar{v} = \lambda$. Then $h(w_A) = aav_iv_ibav_iv_ib$ and $\#_)(w_A) = 2^{i+2} - 1$.

(1c) $v' = \lambda$. Then $h(w_A) = aav_iv_ib$ and $\#_)(w_A) = 2^{i+1}$.

(II) The occurrence of $b$ corresponding to the first occurrence of $a$ in $h(w_A)$ is inside $h(w_A)$. We then have that $h(w_A) = av_iv_ibv'b$ for some $0 \leqslant i \leqslant q$ and $v' \in \{a, b\}^*$, or $h(w_a) = av_iv_ib$ and $\#_)(w_a) = 2^{i+1} - 1$. The former situation is quite analogous to the situation in (1).

Let $(\alpha_2, \alpha_3, \ldots, \alpha_q)$ be a $q - 1$-tuple of functions mapping the words $w_j$ into integers. $\alpha_k(w_i)$ is the number of occurrences of symbols in $w_i$ which are of degree $k$.

We have then

$$\alpha_k(w_0) = 1 \quad \text{for } k = q \quad \text{and} \quad 0 \text{ otherwise, and} \quad \alpha_k(w_n) = 0 \quad \text{for all} \quad k.$$

Assume that $k > \bar{m}$ and $\alpha_k(w_{i+1}) = \alpha_k(w_i) - p$ for some $0 \leqslant i \leqslant q - 1$ and $p > 0$. Then

$$\sum_{r=1}^{\bar{m}} \alpha_{k-r}(w_{i+1}) \geqslant \sum_{r=1}^{\bar{m}} \alpha_{k-r}(w_i) + 2p.$$

To see that we observe that in $i + 1$'th step in the derivation (1) a nonterminal $A$ of degree $k$ must be rewritten as $\alpha$ containing no nonterminal of degree $k$, but containing at least two occurrences of nonterminals of degree at least $k - \bar{m}$ and at most $k - 1$. This is because $\alpha$ has to generate at least $2^k - 2$ right parentheses and that the lenght of $\alpha$ is at most $m$.

Finally we then conclude that there must exist an integer $0 \leqslant i \leqslant n$ such that

$$\sum_{r=0}^{(t+1)\bar{m}} \alpha_{q-r}(w_i) \geqslant t + 1.$$

This means that there in $w_i$ is more than $t$ occurrences of nonterminals generating at least 2 right parentheses in $w$. This contradicts the fact that there are only $t$ nonterminals and that the number of 0's between left and right parentheses all are different in the word $w$.

We have then proved that $L$ is not a parallel context-free language.

THEOREM 2. *The family of context-free language of finite index is contained in the family of parallel context-free languages.*

*Proof.* Let $L$ be a context-free language of index $k$ and $G = (V, T, P, S)$ a context-free grammar of index $k$ generating $G$.

Define $G' = (V', T, P', S')$ to be the grammar where

$$V' = \{A[i] \mid A \in V,\ 1 \leqslant i \leqslant k\},$$

$P'$ contains all productions of the form

$$A[i_0] \to x_1 A_1[i_1]\ x_2 A_2[i_2] \cdots A_n[i_n]\ x_{n+1}$$

where $x_i \in T^*$ for $1 \leqslant i \leqslant n+1$, $1 \leqslant i_j \leqslant k$ for all $0 \leqslant j \leqslant n$, and $A \to x_1 A_1 x_2 A_2 \cdots A_n x_{n+1}$ is a production in $P$.

$$S' = S[1].$$

It is now obvious that $L(G) = L(G')$, and that $G'$ is of index $k$. This means that we can choose derivations such that no nonterminal in $V'$ occurs more than once in every sentential form in these derivations. It is then clear that $G'$ generates $L$ also in parallel.

RECEIVED: June 20, 1974

## REFERENCES

GINSBURG, S. AND SPANIER, E. (1968), Derivation-bounded languages, *J. Comput. System Sci.* **2**, 228.

ROSENFELD, A. (1971), Isotonic grammars, parallel grammars, and picture grammars, *Mach. Intelligence* **6**, 281.

SALOMAA, A. (1973), "Formal Languages," Academic Press, New York.

SIROMONEY, R. AND KRITHIVASAN, K. (1974), Parallel context-free languages, *Inform. Contr.* **24**, 155.